



Large deviations for the Fleming–Viot process with neutral mutation and selection¹

Donald A. Dawson^{a,*}, Shui Feng^b

^a *The Fields Institute, 222 College Street, Toronto, Ont., Canada M5T 3J1*

^b *Department of Mathematics and Statistics, McMaster University, Hamilton, Ont., Canada L8S 4K1*

Received 8 July 1997; accepted 27 April 1998

Abstract

Large deviation principles are established for the Fleming–Viot processes with neutral mutation and selection, and the corresponding equilibrium measures as the sampling rate goes to 0. All results are first proved for the finite allele model, and then generalized, through the projective limit technique, to the infinite allele model. Explicit expressions are obtained for the rate functions. © 1998 Elsevier Science B.V. All rights reserved.

AMS classification: primary 60F10; secondary 92D10

Keywords: Fleming–Viot process; Large deviations; Relative entropy

1. Introduction

The Fleming–Viot process is a measure-valued process describing the evolution of the distribution of genotypes in a population. In the case of two alleles it reduces to a one-dimensional diffusion process that approximates the classical Wright–Fisher model. The standard model in population genetics involves mutation, replacement sampling, and selective advantages among various genotypes (see Hartl and Clark (1989)).

Let E be a compact metric space, $C(E)$ be the set of continuous functions on E , and $M_1(E)$ denote the space of all probability measures on E with the topology of weak convergence. Let A be the generator of a Markov process on E with domain $D(A)$. Define $\mathcal{D} = \{F : F(\mu) = f(\langle \mu, \phi \rangle), f \in C_b^\infty(R), \phi \in D(A), \mu \in M_1(E)\}$. Then the generator of the Fleming–Viot process is

$$\begin{aligned}\mathcal{L}F(\mu) &= \int_E \left(A \frac{\delta F(\mu)}{\delta \mu(x)} \right) \mu(dx) + \frac{\gamma}{2} \int_E \int_E \left(\frac{\delta^2 F(\mu)}{\delta \mu(x) \delta \mu(y)} \right) Q(\mu; dx, dy) \\ &= f'(\langle \mu, \phi \rangle) \langle \mu, A\phi \rangle + \frac{\gamma}{2} \int \int \phi(x)\phi(y) Q(\mu; dx, dy),\end{aligned}\tag{1.1}$$

* Corresponding author. E-mail: don@fields.utoronto.ca.

¹ Research supported by the Natural Science and Engineering Research Council of Canada.

where

$$\begin{aligned}\delta F(\mu)/\delta\mu(x) &= \lim_{\varepsilon\rightarrow 0+} \varepsilon^{-1}\{F(\mu+\varepsilon\delta_x)-F(\mu)\}, \\ \delta^2 F(\mu)/\delta\mu(x)\delta\mu(y) &= \lim_{\varepsilon_1\rightarrow 0+, \varepsilon_2\rightarrow 0+} (\varepsilon_1\varepsilon_2)^{-1}\{F(\mu+\varepsilon_1\delta_x+\varepsilon_2\delta_y)-F(\mu)\}, \\ Q(\mu;dx,dy) &= \mu(dx)\delta_x(dy)-\mu(dx)\mu(dy),\end{aligned}$$

and δ_x stands for the Dirac measure at $x\in E$. The domain of \mathcal{L} is \mathcal{D} . E is called the type space, A is known as the mutation operator, and the last term in Eq. (1.1) describes the continuous sampling. If the mutation operator has the form of $Af(x)=\frac{\theta}{2}\int(f(y)-f(x))v_0(dy)$ with $v_0\in M_1(E)$, we call the process a Fleming–Viot process with neutral mutation. It is known that the Fleming–Viot process with neutral mutation has a unique reversible probability measure (cf. Ethier and Kurtz, 1993).

In the present article we will consider the limiting behavior of this process as $\gamma\rightarrow 0$. In the first principal result we establish a large deviation principle (henceforth, LDP) for the sequence of reversible measures. It turns out that the sequence converges to the probability fixed point v_0 of the mutation operator exponentially fast and for $\theta=1, v\ll v_0$ the deviation is characterized by the relative entropy defined as

$$H(v_0|v)=\begin{cases}\int_E h\log h\,dv & \text{if } v_0\ll v, \\ \infty & \text{otherwise,}\end{cases}\tag{1.2}$$

where h is the Radon–Nikodym derivative of v_0 with respect to v . It is known, by Sanov’s theorem, that the empirical measure $\{\frac{1}{n}\sum_{k=1}^n\delta_{X_k}\}_{n\geq 1}$ of an i.i.d. sequence of random variables with common distribution v_0 converges exponentially fast to v_0 as n goes to infinity, and the deviation is characterized by the relative entropy $H(v|v_0)$. Our example here may be the first among the large deviation literature that has this “reversed” form of relative entropy as rate function. In Sanov’s theorem, the influence of sampling is dominant while in the Fleming–Viot case this influence decreases to zero. This may be an explanation for the “reversed” expression of the two rate functions.

The second-principal result of this article establishes a path level LDP for the Fleming–Viot process with neutral mutation and with selection. This can be viewed as an Freidlin–Wentzell-type result in infinite dimension. Furthermore, the existing results on large deviations for finite-dimensional diffusions usually assume either the diffusion coefficient is non-degenerate or the square root of the diffusion coefficient is uniformly Lipschitz continuous, but our model with finite alleles satisfies neither of them. Hence our results also include an extension of the finite-dimensional Freidlin–Wentzell theory.

The sampling rate γ can be interpreted as the inverse population size and the large deviation results of this paper describe the deviations from the “infinite population” deterministic limit.

The large deviation result for equilibrium measures is proved in Section 2 and the path level LDP is proved in Section 3. We will first prove the LDP for the Fleming–Viot process with neutral mutation, and then, by the Cameron–Martin–Girsanov transformation, for the case with selection. For processes without selection we will first prove the result for the finite allele model, and then generalize, through the projective limit technique, to the infinite allele model.

2. LDP for equilibrium measures

Let the following objects be given: X a Hausdorff topological space, \mathcal{B}_X a σ -algebra of space X , $\{P_\varepsilon\}_{\varepsilon>0}$ a family of probability measures on (X, \mathcal{B}_X) , $\{a_\varepsilon\}_{\varepsilon>0}$ a family of positive numbers tending to zero as ε goes to zero, and a function $I: X \rightarrow [0, \infty]$.

Definition 2.1. The function $I: X \rightarrow [0, \infty]$ is called a rate function if it is lower semicontinuous. A rate function is called the good rate function if for any $r \geq 0$, the level set $\Phi_I(r) = \{x \in X: I(x) \leq r\}$ is compact. The constant a_ε is called the speed.

Definition 2.2. $\{P_\varepsilon\}$ satisfies a LDP with the rate function (or good rate function) I if

1. for each \mathcal{B}_X -measurable open subset G of X

$$\liminf_{\varepsilon \rightarrow 0} a_\varepsilon \log P_\varepsilon(G) \geq - \inf_{x \in G} I(x); \quad (2.1)$$

2. for each \mathcal{B}_X -measurable closed subset B of X

$$\limsup_{\varepsilon \rightarrow 0} a_\varepsilon \log P_\varepsilon(B) \leq - \inf_{x \in B} I(x); \quad (2.2)$$

Remark. The \mathcal{B}_X -measurable condition is needed when \mathcal{B} is not the Borel σ -algebra and not all open or closed sets are \mathcal{B}_X -measurable. This situation may occur when the space X is not separable. When the space X is compact, all rate functions are good rate functions. Also the function I in the above definition is unique when X is a regular topological space. For an excellent introduction to basic concepts and techniques of large deviations refer to Dembo and Zeitouni (1993).

In this section we will establish LDP for the equilibrium measures of the Fleming–Viot processes with neutral mutation and with selection. We will start with the case when the type space E is finite. Then, by using the projective limit technique, we obtain results for the case of $E = [0, 1]$.

For any $n \geq 1$, let $E = \{1, 2, \dots, n\}$. The space $M_1(E)$, the set of all probability measures on E , can be identified with the $(n-1)$ -dimensional simplex

$$\Delta_n = \left\{ x = (x_1, \dots, x_n): x_i \geq 0, \quad i = 1, \dots, n; \quad \sum_{i=1}^n x_i = 1 \right\}.$$

The Fleming–Viot process with neutral mutation reduces to the neutral one-locus n -allele diffusion process with generator

$$\mathcal{A}_\gamma = \frac{\gamma}{2} \sum_{i,j=1}^n x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^n \left(\sum_{j=1}^n x_j q_{ji} \right) \frac{\partial}{\partial x_i},$$

where q_{ji} ($j \neq i$) is the intensity of a mutation from allele j to allele i , and $q_{jj} = -\sum_{i \neq j} q_{ji}$.

Let m denote the Lebesgue measure on Δ_n . If the infinitesimal matrix (q_{ij}) is irreducible, then this diffusion has a unique stationary distribution which is absolutely continuous with respect to m . (cf. Shiga, 1981).

In the special case of parent-independent mutation, i.e.,

$$q_{ij} = \frac{\theta}{2} p_j > 0, \quad 1 \leq i \neq j \leq n, \quad \sum_{i=1}^n p_i = 1, \quad \theta > 0.$$

Wright (1949) discovered that for $p = (p_1, \dots, p_n)$, the unique stationary distribution $\Pi_{\theta, \gamma, p} \in M_1(\Delta_n)$ is the Dirichlet distribution with parameters p_1, \dots, p_n given by

$$\Pi_{\theta, \gamma, p}(dx) = \frac{\Gamma[\gamma^{-1}\theta]}{\Gamma(\gamma^{-1}\theta p_1) \cdots \Gamma(\gamma^{-1}\theta p_n)} x_1^{\gamma^{-1}\theta p_1 - 1} \cdots x_n^{\gamma^{-1}\theta p_n - 1} dx_1 \cdots dx_{n-1}.$$

For any Borel measurable subset C of Δ_n , we have by Stirling's formula

$$\begin{aligned} \log \Pi_{\theta, \gamma, p}(C) &= \log \left\{ \frac{\sqrt{2\pi}(\gamma^{-1}\theta)^{\gamma^{-1}\theta - 1/2} e^{\alpha\gamma/12\theta}}{(\sqrt{2\pi})^n (\gamma^{-1}\theta p_1)^{\gamma^{-1}\theta p_1 - 1/2} e^{\alpha_1\gamma/12\theta} p_1 \cdots (\gamma^{-1}\theta p_n)^{\gamma^{-1}\theta p_n - 1/2} e^{\alpha_n\gamma/12\theta} p_n} \right. \\ &\quad \times \left. \int_C x_1^{\gamma^{-1}\theta p_1 - 1} \cdots x_n^{\gamma^{-1}\theta p_n - 1} dx_1 \cdots dx_{n-1} \right\} \\ &= \frac{n-1}{2} \log \frac{1}{2\gamma\pi} + \frac{1}{2} \log \frac{(\theta p_1) \cdots (\theta p_n)}{\theta} + \frac{\gamma}{12\theta} \left[\alpha - \frac{\alpha_1}{p_1} - \cdots - \frac{\alpha_n}{p_n} \right] \\ &\quad - \gamma^{-1}\theta \sum_{i=1}^n p_i \log p_i + \log \int_C x_1^{\gamma^{-1}\theta p_1 - 1} \cdots x_n^{\gamma^{-1}\theta p_n - 1} dx_1 \cdots dx_{n-1}, \end{aligned} \quad (2.3)$$

where $0 < \alpha, \alpha_1, \dots, \alpha_n < 1$ are some constants.

For any $\varepsilon > 0$, let $C_\varepsilon = \{x \in C: \min_{1 \leq i \leq n} x_i \geq \varepsilon\}$. For any measurable function f on Δ_n , $\|f\|_{L^{\gamma^{-1}}}$ denotes the $L^{\gamma^{-1}}$ norm of f with respect to measure m . Then we have for $\gamma < \min_{1 \leq i \leq n} \{\theta p_i\}$,

$$\begin{aligned} \|I_C e^{\sum_{i=1}^n \theta p_i \log x_i}\|_{L^{\gamma^{-1}}} &= \left[\int_C x_1^{\gamma^{-1}\theta p_1} \cdots x_n^{\gamma^{-1}\theta p_n} dx_1 \cdots dx_{n-1} \right]^\gamma \\ &\leq \left[\int_C x_1^{\gamma^{-1}\theta p_1 - 1} \cdots x_n^{\gamma^{-1}\theta p_n - 1} dx_1 \cdots dx_{n-1} \right]^\gamma \\ &= \|I_C e^{\sum_{i=1}^n (\theta p_i - \gamma) \log x_i}\|_{L^{\gamma^{-1}}} \\ &\leq \|I_{A_\varepsilon} e^{\sum_{i=1}^n \theta p_i \log x_i}\|_{L^{\gamma^{-1}}} \left(\frac{1}{\varepsilon} \right)^{n\gamma} + m(C \setminus C_\varepsilon) \\ &\leq \|I_C e^{\sum_{i=1}^n \theta p_i \log x_i}\|_{L^{\gamma^{-1}}} \left(\frac{1}{\varepsilon} \right)^\gamma + m(C \setminus C_\varepsilon). \end{aligned} \quad (2.4)$$

Letting $\gamma \rightarrow 0$, then $\varepsilon \rightarrow 0$, we get

$$\begin{aligned} \lim_{\gamma \rightarrow 0} \left[\int_C x_1^{\gamma^{-1}\theta p_1 - 1} \cdots x_n^{\gamma^{-1}\theta p_n - 1} dx_1 \cdots dx_{n-1} \right]^\gamma \\ = \text{ess sup} \left\{ I_C e^{\sum_{i=1}^n \theta p_i \log x_i} : x \in \Delta_n \right\}. \end{aligned} \quad (2.5)$$

For any subset B of Δ_n , we have

$$\text{ess sup} \{I_B e^{\sum_{i=1}^n \theta p_i \log x_i} : x \in \Delta_n\} \leq e^{-\inf_{x \in B} \sum_{i=1}^n \theta p_i \log 1/x_i}. \quad (2.6)$$

On the other hand, for any open subset G of Δ_n ,

$$\begin{aligned} \text{ess sup}\{I_G e^{\sum_{i=1}^n \theta p_i \log x_i} : x \in \Delta_n\} &= \text{ess sup}\{e^{\sum_{i=1}^n \theta p_i \log x_i} : x \in G\} \\ &= e^{-\inf_{x \in G} \sum_{i=1}^n \theta p_i \log 1/x_i}. \end{aligned} \quad (2.7)$$

By Eqs. (2.3), (2.5) and (2.6) we get that for any closed subset B of Δ_n

$$\limsup_{\gamma \rightarrow 0} \gamma \log \Pi_{\theta, \gamma, p}(B) \leq -\theta \inf_{x \in B} \left\{ \sum_{i=1}^n \left(p_i \log \frac{1}{x_i} + p_i \log p_i \right) \right\}. \quad (2.8)$$

From Eqs. (2.3) (2.5) and (2.7), we obtain that for any open subset G of Δ_n

$$\liminf_{\gamma \rightarrow 0} \gamma \log \Pi_{\theta, \gamma, p}(G) \geq -\theta \inf_{x \in G} \left\{ \sum_{i=1}^n \left(p_i \log \frac{1}{x_i} + p_i \log p_i \right) \right\}. \quad (2.9)$$

Note that in the present situation the relative entropy $H(p|x)$ of p with respect to x is given by

$$H(p|x) = \sum_{i=1}^n p_i \log \frac{p_i}{x_i},$$

and is non-negative, continuous. This combined with the compactness of Δ_n implies that all level sets of $\{x \in \Delta_n : H(p|x) \leq r\}$ are compact. Thus, we have proved the following theorem.

Theorem 2.1. *When the mutation is parent independent, the family $\{\Pi_{\theta, \gamma, p}\}_{\gamma > 0}$ satisfies a LDP on space Δ_n with the good rate function $I_p^\theta(x) = \theta H(p|x)$ as γ goes to zero.*

Remark. It is well known that the relative entropy $H(x|p)$ is the rate function describing the large deviations of the empirical measure of an i.i.d. sequence of random variables with common distribution p . In the case of $\theta = 1$ the rate function in Theorem 2.1 has an “reversed” expression $H(p|x)$.

In Theorem 2.1, the probability measure p is positive for every type i in E . By making $x_i = 0$ whenever $p_i = 0$, the Dirichlet distribution can be defined for any $p \in \Delta_n$ as follows. Without loss of generality, let $p = (p_1, \dots, p_r, 0, \dots, 0)$, $r < n$, $\tilde{p} = (p_1, \dots, p_r)$. We define

$$\Pi_{\theta, \gamma, p} = \Pi_{\theta, \gamma, \tilde{p}} \times \delta_0^{\otimes n-r}. \quad (2.10)$$

For any $p, x \in \Delta_n$, define

$$I_p^\theta(x) = \begin{cases} \theta H(p|x) & \text{if } x \ll p, \\ \infty & \text{otherwise,} \end{cases} \quad (2.11)$$

Then a generalized version of Theorem 2.1 can be stated as follows.

Theorem 2.2 *For any $p \in \Delta_n$, the family $\{\Pi_{0,\gamma,p}\}_{\gamma>0}$ satisfies a LDP on space Δ_n with the good rate function $I_p^\theta(x)$ as γ goes to zero.*

Next, we consider the case when the type space $E = [0, 1]$. Let $M_1([0, 1])$ be the collection of all probability measures on $[0, 1]$ with the topology of weak convergence. Another topology on $M_1([0, 1])$ needed in the sequel is called the τ topology which is the smallest topology such that for each bounded measurable function f on $[0, 1]$, the map $\mu \rightarrow \int_0^1 f(z)\mu(dz)$ is continuous. The τ -topology on $M_1([0, 1])$ is stronger than the weak topology. But when the type space E is finite, the two topologies coincide with each other. We use $M_1^\tau([0, 1])$ to denote the space of all probability measures on $[0, 1]$ equipped with the τ topology.

Note that for each $x \in [0, 1]$, let $V_x = \{\mu \in M_1([0, 1]): \mu(x) > \frac{3}{4}\}$. Then each V_x is a non-empty open set in the τ topology and for different $x, y \in [0, 1]$, V_x and V_y are disjoint. Since there are uncountable number of such open sets, the space $M_1^\tau([0, 1])$ is not separable. The σ -algebra \mathcal{B}^τ on $M_1^\tau([0, 1])$ is defined to be the smallest σ -algebra such that for every bounded, measurable function f on $[0, 1]$, the map $\mu \rightarrow \int_0^1 f(x)d\mu(x)$ is measurable. It is known (cf. Dembo and Zeitouni (1993)) that the Borel σ -algebra of space $M_1([0, 1])$ is the same as \mathcal{B}^τ . We use \mathcal{B} to denote this common σ -algebra throughout the remainder of this section.

Let

$$\mathcal{P} = \{\{B_1, \dots, B_r\}: r \geq 1, B_1, \dots, B_r \text{ is a partition of } [0, 1] \text{ by Borel measurable sets}\} \tag{2.12}$$

be the collection of all finite partitions of $[0, 1]$. For any $\mu \in M_1([0, 1])$, $j = \{B_1, \dots, B_r\} \in \mathcal{P}$, define $\pi_j(\mu) = (\mu(B_1), \dots, \mu(B_r))$.

For any $\mu, \nu \in M_1([0, 1])$, let $H(\mu|\nu)$ be the relative entropy of μ with respect to ν defined in Eq. (1.2). Then it is known (cf. Donsker and Varadhan, 1975) that

$$\begin{aligned} H(\mu|\nu) &= \sup_{g \in C([0,1])} \left\{ \int_0^1 g d\mu - \log \int_0^1 e^g d\nu \right\} \\ &= \sup_{g \in B([0,1])} \left\{ \int_0^1 g d\mu - \log \int_0^1 e^g d\nu \right\}, \end{aligned} \tag{2.13}$$

where $C([0, 1]), B([0, 1])$ are the sets of continuous functions and bounded measurable functions on $[0, 1]$, respectively.

The following expression of the relative entropy will be repeatedly used in the sequel.

Lemma 2.3. *For any $\mu, \nu \in M_1([0, 1])$,*

$$H(\mu|\nu) = \sup_{j \in \mathcal{P}} H(\pi_j(\mu)|\pi_j(\nu)). \tag{2.14}$$

Proof. If μ is not absolutely continuous with respect to ν , then both sides of Eq. (2.14) are infinity. Now, we assume that $h = d\mu/d\nu$ exists. Then for any $j = (B_1, \dots, B_r) \in \mathcal{P}$,

choose a function $g \in B([0, 1])$ such that

$$g(z) = \sum_{i: \mu(B_i) > 0} \log \frac{\mu(B_i)}{\nu(B_i)} I_{B_i}(z).$$

By direct calculation we have

$$H(\pi_j(\mu) | \pi_j(\nu)) = \int_0^1 g(z) d\mu - \log \int_0^1 e^{g(z)} d\nu.$$

By Eq. (2.13), we get $H(\pi_j(\mu) | \pi_j(\nu)) \leq H(\mu | \nu)$ which implies that $\sup_{j \in \mathcal{P}} H(\pi_j(\mu) | \pi_j(\nu)) \leq H(\mu | \nu)$. On the other hand, for any $n \geq 1$, let

$$h_n(x) = \sum_{k=1}^{n2^n} (k-1)/2^n I_{B_k}(x) + n I_{A_n}(x),$$

where $B_k = \{z: (k-1)/2^n \leq h(z) < k/2^n\}$, $A_n = \{z: h(z) \geq n\}$. The fact that h is integrable with respect to ν implies that h_n is an increasing sequence of nonnegative simple functions converging to h almost surely with respect to ν . Let $\ell = \{B_1, \dots, B_{n2^n}, A_n\}$. Then we have

$$\sup_{j \in \mathcal{P}} H(\pi_j(\mu) | \pi_j(\nu)) \geq H(\pi_\ell(\mu) | \pi_\ell(\nu)) \geq \int_0^1 h_n \log h_n d\nu.$$

Let n go to infinity, we get Eq. (2.14) by Eq. (1.2) and the monotone convergence theorem. \square

Now, we are ready to prove large deviation results for equilibrium measures of some Fleming–Viot processes. A Fleming–Viot process with neutral mutation has a generator of the following form:

$$\mathcal{L}F(\mu) = \int_E \left(A \frac{\delta F(\mu)}{\delta \mu(z)} \right) \mu(dz) + \frac{\gamma}{2} \int_E \int_E \left(\frac{\delta^2 F(\mu)}{\delta \mu(z) \delta \mu(y)} \right) Q(\mu; dz, dy), \quad (2.15)$$

where $E = [0, 1]$, $\theta > 0$, $\nu_0 \in M_1([0, 1])$, and

$$Af(z) = \frac{\theta}{2} \int_0^1 (f(y) - f(z)) \nu_0(dy),$$

$$Q(\mu; dz, dy) = \mu(dz) \delta_z(dy) - \mu(dz) \mu(dy).$$

For any symmetric bounded measurable function $V(z, y) \in B([0, 1] \times [0, 1])$, let

$$V(\mu) = \int_0^1 \int_0^1 V(z, y) \mu(dz) \mu(dy),$$

and

$$\left\langle \frac{\delta V(\mu)}{\delta \mu}, \frac{\delta F}{\delta \mu} \right\rangle = \int_0^1 \int_0^1 \int_0^1 \frac{\delta F}{\delta \mu}(z) [V(z, y) - V(y, w)] \mu(dz) \mu(dy) \mu(dw).$$

Then the generator of a Fleming–Viot process with neutral mutation and selection takes the form:

$$\mathcal{L}_V F(\mu) = \mathcal{L} F(\mu) + \left\langle \frac{\delta V(\mu)}{\delta \mu}, \frac{\delta F}{\delta \mu} \right\rangle, \tag{2.16}$$

where V is called the fitness function.

It is known (cf. Theorems 3.1 and 3.2 in Ethier and Kurtz, 1993) that the martingale problem associated with generators \mathcal{L} and \mathcal{L}_V are well-posed. Shiga (1990) shows that the Fleming–Viot process with generator \mathcal{L} has a unique, reversible stationary distribution $\Pi_{\theta, \gamma, v_0} \in M_1(M_1([0, 1]))$, which is the distribution of a $M_1([0, 1])$ -valued random variable v characterized by the property that whenever $r \geq 2$ and B_1, \dots, B_r is a partition of $[0, 1]$, $(v(B_1), \dots, v(B_r))$ has the Dirichlet distribution with parameters $(\theta/\gamma)v_0(B_1), \dots, (\theta/\gamma)v_0(B_r)$. (Note that zero parameters can be removed to create a Dirichlet distribution on a simplex with dimension less than r .) The Fleming–Viot process with neutral mutation and selection also has a unique, reversible stationary distribution given by

$$\Pi_{\theta, \gamma, v_0, V}(d\mu) = Z^{-1} \exp \left[\frac{V(\mu)}{\gamma} \right] \Pi_{\theta, \gamma, v_0}(d\mu), \tag{2.17}$$

where Z is the normalizing constant (cf. Ethier and Kurtz, 1994).

Theorem 2.4. *The family $\{\Pi_{\theta, \gamma, v_0}\}$ satisfies a LDP on space $M_1^r([0, 1])$ with the good rate function*

$$I_{v_0}^\theta(\mu) = \begin{cases} \theta H(v_0 | \mu) & \text{if } \mu \ll v_0, \\ \infty & \text{otherwise.} \end{cases} \tag{2.18}$$

Proof. First note that the set of all finite partitions \mathcal{P} , partially ordered by $j \succ i$ iff j is finer than i , is a partially ordered right-filtering set. For every $j = (B_1, \dots, B_r) \in \mathcal{P}$, let

$$\mathcal{X}_j = \left\{ x = (x_{B_1}, \dots, x_{B_r}) : x_{B_i} \geq 0, \ i = 1, \dots, r; \sum_{i=1}^r x_{B_i} = 1 \right\}.$$

For any $i = (C_1, \dots, C_l)$, $j = (B_1, \dots, B_r) \in \mathcal{P}$, $j \succ i$, define

$$\pi_{ij} : \mathcal{X}_j \rightarrow \mathcal{X}_i, (x_{B_1}, \dots, x_{B_r}) \rightarrow \left(\sum_{B_k \subset C_1} x_{B_k}, \dots, \sum_{B_k \subset C_l} x_{B_k} \right).$$

Let \mathcal{X} represent the projective limit of the family $\{\mathcal{X}_j, \pi_{ij}, i, j \in J, j \succ i\}$, π_j be the projective mapping. Then we have $M_1([0, 1]) \subset \mathcal{X}$ by identifying $\mu \in M_1([0, 1])$ with the projective limit of $\{(\mu(B_1), \dots, \mu(B_r)) : j = (B_1, \dots, B_r) \in \mathcal{P}\}$.

On the other hand, any element μ of \mathcal{X} can be viewed as a finitely additive measure on $[0, 1]$. For any $f \in C([0, 1])$ and $\mu \in \mathcal{X}$, we define the following “abstract integral” of f with respect to μ :

$$\langle \mu, f \rangle = \lim_{n \rightarrow \infty} \sum_{k=0}^{2^n-1} f(2^{-n}k) x_{B_k},$$

where $B_0 = [0, 2^{-n})$, $B_k = (2^{-n}k, 2^{-n}(k+1)]$ for $k = 1, \dots, 2^{-n} - 1$. The existence of this limit is guaranteed since $\sum_{k=0}^{2^n-1} f(2^{-n}k)x_{B_k}$ is a bounded Cauchy sequence. From this definition and the fact that any decreasing sequence $f_n \in C([0, 1])$ which converges to zero pointwisely converges to zero uniformly (Dini's theorem), we conclude that $\langle \mu, f \rangle$ is a linear form on space $C([0, 1])$ satisfying:

- (a) $\langle \mu, f \rangle \geq 0$ for $f \geq 0$;
- (b) $\langle \mu, f \rangle = 1$ for $f \equiv 1$;
- (c) $\langle \mu, f \rangle \rightarrow 0$ as $f \downarrow 0$ pointwisely.

By the Daniell–Stone theorem (see e.g. p. 197 of Bauer, 1981), μ is a probability measure on the Borel σ -algebra of $[0, 1]$. The projective topology obtained is just the τ -topology. Hence we have $\mathcal{X} = M_1^\tau([0, 1])$. By Theorems 2.2 and 3.3 of Dawson and Gärtner (1987) we have that the family $\{\Pi_{\theta, \gamma, v_0}\}$ satisfies a LDP on space \mathcal{X} with the good rate function

$$I(\mu) = \sup \{I_{\pi_j(v_0)}^\theta(\pi_j(\mu)) : j \in \mathcal{P}\}. \quad (2.19)$$

From Eq. (2.3), we see that $I(\mu) = I_{v_0}^\theta(\mu)$. \square

Now, let

$$C(\theta, v_0, V) = \sup_{\mu \in M_1(E)} \{V(\mu) - I_{v_0}^\theta(\mu)\}, \quad I_{v_0}^{\theta, V}(\mu) = C(\theta, v_0, V) - [V(\mu) - I_{v_0}^\theta(\mu)].$$

Then we have the following LDP.

Theorem 2.5. *The family $\{\Pi_{\theta, \gamma, v_0, V}\}$ satisfies a LDP on space $M_1([0, 1])$ with the good rate function $I_{v_0}^{\theta, V}(\mu)$.*

Proof. Since $C(\theta, v_0, V)$ is a constant, $V(\mu)$ is bounded continuous in the τ -topology, any level set of function $I_{v_0}^{\theta, V}(\mu)$ is a τ -closed subset of a level set of function $I_{v_0}^\theta(\mu)$. Hence $I_{v_0}^{\theta, V}(\mu)$ is a good rate function in the τ -topology, and thus in the weak topology too.

By using Varadhan's Lemma, we have for any $F(\mu) \in C(M_1([0, 1]))$,

$$\begin{aligned} \lim_{\gamma \rightarrow 0} \gamma \log Z &= \lim_{\gamma \rightarrow 0} \gamma \log \int_{M_1([0, 1])} e^{V(\mu)/\gamma} \Pi_{\theta, \gamma, v_0}(\mathrm{d}\mu) = C(\theta, v_0, V), \\ \lim_{\gamma \rightarrow 0} \gamma \log \int_{M_1([0, 1])} e^{F(\mu)/\gamma} \Pi_{\theta, \gamma, v_0, V}(\mathrm{d}\mu) &= C(\theta, v_0, F + V) - C(\theta, v_0, V). \end{aligned}$$

Since $M_1([0, 1])$ is compact and thus the family $\Pi_{\theta, \gamma, v_0, V}$ is exponential tight, by Bryc's inverse Varadhan Lemma (cf. Section 4.4 of Dembo and Zeitouni (1993)), we get that the family $\{\Pi_{\theta, \gamma, v_0, V}\}$ satisfies a LDP on space $M_1([0, 1])$ with the good rate function $I_{v_0}^{\theta, V}(\mu)$. \square

3. Path level LDP

In this section we will establish three LDPs at the path level: LDP for finite type (or allele) model, LDP for the Fleming–Viot process with neutral mutation, and LDP for the Fleming–Viot process with neutral mutation and selection. The novelty of our result for finite type model is that the diffusion coefficient of the corresponding diffusion process is degenerate and the square root of the diffusion coefficient is non-Lipschitz. This is an extension of the Freidlin–Wentzell theory.

3.1. LDP For finite allele model

Define

$$\mathcal{S}_n = \left\{ x = (x_1, \dots, x_{n-1}) : x_i \geq 0, \ i = 1, \dots, n-1; \sum_{i=1}^{n-1} x_i \leq 1 \right\}.$$

The Fleming–Viot process with finite allele and neutral mutation is a finite-dimensional diffusion process described by the following system of stochastic differential equations:

$$dx_k^\varepsilon(t) = b_k(x^\varepsilon(t)) dt + \varepsilon \sum_{l=1}^{n-1} \sigma_{kl}(x^\varepsilon(t)) dB_l(t), \quad 1 \leq k \leq n-1, \quad (3.1)$$

where $x^\varepsilon(t) = (x_1^\varepsilon(t), \dots, x_{n-1}^\varepsilon(t))$, $b_k(x^\varepsilon(t)) = \frac{\theta}{2}(p_k - x_k^\varepsilon(t))$, and $\sigma(x^\varepsilon(t)) = (\sigma_{kl}(x^\varepsilon(t)))_{1 \leq k, l \leq n-1}$ is given by

$$\sigma(x^\varepsilon(t))\sigma'(x^\varepsilon(t)) = D(x^\varepsilon(t)) = (x_k^\varepsilon(t)(\delta_{kl} - x_l^\varepsilon(t)))_{1 \leq k, l \leq n-1},$$

and $B_l(t)$, $1 \leq l \leq n-1$ are independent Brownian motions, $\gamma = \varepsilon^2$, $p_k = v_0(k) > 0$.

For a fixed $T > 0$ and $x \in \mathcal{S}_n$, let $C([0, T], \mathcal{S}_n)$ be the space of all \mathcal{S}_n -valued continuous functions on $[0, T]$ endowed with the uniform topology, and P_x^ε denote the law of $x^\varepsilon(\cdot)$ starting at x .

Lemma 3.1. *The family $\{P_x^\varepsilon\}_{\varepsilon > 0}$ is exponentially tight on $C([0, T], \mathcal{S}_n)$ for all $x \in \mathcal{S}_n$.*

Proof. Let $c_1 \geq \sup_{x \in \mathcal{S}_n} \|b(x)\|$, $c_2 = \sup_{\|\theta\|=1, x \in \mathcal{S}_n} \langle \theta, D(x)\theta \rangle > 0$. For any $a > 0$, choose $m_0 \geq 1$, $\varepsilon_0 > 0$ such that for $m \geq m_0$, $\varepsilon \leq \varepsilon_0$, we have $2(n-1)m^2T \geq 1$ and $a\varepsilon^{-2} > 1$. Define

$$K_a = \bigcap_{m \geq m_0} \left\{ x(\cdot) \in C([0, T], \mathcal{S}_n); \sup_{t, s \in [0, T], |t-s| < \gamma_m} |x(t) - x(s)| \leq \rho_m \right\},$$

where

$$\gamma_m = \frac{1}{m^2}, \quad \rho_m = c_1\gamma_m + \left\{ 2c_2(n-1)a \left[\frac{1}{m} + \gamma_m \ln \frac{2(n-1)T}{\gamma_m} \right] \right\}^{1/2}.$$

Let $b(x) = (b_1(x), \dots, b_{n-1}(x))$. Then we have

$$\begin{aligned}
 P_x^\varepsilon \{x(\cdot) \in K_a^c\} &\leq \sum_{m=m_0}^{\infty} P \left\{ \sup_{t,s \in [0,T], |t-s| < \gamma_m} |x^\varepsilon(t) - x^\varepsilon(s)| > \rho_m \right\} \\
 &\leq \sum_{m=m_0}^{\infty} P \left\{ \sup_{t,s \in [0,T], |t-s| < \gamma_m} |x^\varepsilon(t) - x^\varepsilon(s)| \right. \\
 &\quad \left. - \left\{ \int_s^t b(x^\varepsilon(\tau)) d\tau \right\} > \rho_m - c_1 \gamma_m \right\} \\
 &\leq \sum_{m=m_0}^{\infty} \frac{2(n-1)T}{\gamma_m} \exp \left(-\frac{(\rho_m - c_1 \gamma_m)^2}{2c_2(n-1)\gamma_m \varepsilon^2} \right) \\
 &= \sum_{m=m_0}^{\infty} \frac{2(n-1)T}{\gamma_m} \exp \left(-\frac{a}{\gamma_m \varepsilon^2} (1/m + \gamma_m \ln(2(n-1)T/\gamma_m)) \right) \\
 &\leq \sum_{m=m_0}^{\infty} \frac{2(n-1)T}{\gamma_m} \left(\frac{\gamma_m}{2(n-1)T} \right)^{a/\varepsilon^2} \\
 &\quad \times \exp(-am/\varepsilon^2) \leq \frac{\exp(-\frac{a}{\varepsilon^2})}{1 - \exp(-\frac{a}{\varepsilon^2})}, \tag{3.2}
 \end{aligned}$$

where the third inequality is obtained by using Theorem 4.2.1 in Stroock and Varadhan (1979). Letting ε go to zero, we get

$$\limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log P_x^\varepsilon \{K_a^c\} \leq -a. \tag{3.3}$$

The lemma follows since K_a is a compact set in $C([0, T], \mathcal{S}_n)$. \square

Let $\partial \mathcal{S}_n$ and \mathcal{S}_n° denote the boundary and interior of \mathcal{S}_n , respectively. By direct calculation we get $\det(A(x)) = x_1 \cdots x_{n-1} (1 - \sum_{i=1}^{n-1} x_i)$. Thus for any $x \in \mathcal{S}_n^\circ$, $D(x)$ is invertible and the inverse is given by

$$D^{-1}(x) = \begin{pmatrix} \frac{1 - \sum_{1 \leq i \leq n-1, i \neq 1} x_i}{x_1 x_n} & \frac{1}{x_n} & \cdots & \frac{1}{x_n} \\ \frac{1}{x_n} & \frac{1 - \sum_{1 \leq i \leq n-1, i \neq 2} x_i}{x_2 x_n} & \cdots & \frac{1}{x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{1}{x_n} & \frac{1}{x_n} & \cdots & \frac{1 - \sum_{1 \leq i \leq n-1, i \neq n-1} x_i}{x_{n-1} x_n} \end{pmatrix} \tag{3.4}$$

where $x_n = 1 - \sum_{i=1}^{n-1} x_i$.

For any x in \mathcal{S}_n and φ in $C([0, T], \mathcal{S}_n)$, let

$$\begin{aligned}
 H_1^x &= \left\{ \varphi: \varphi(t) = x + \int_0^t g(s) ds, g \in L^2([0, T]) \right\}, \\
 K_\varphi &= \left\{ g \in H_1^x: \varphi(t) = x + \int_0^t b(\varphi(s)) ds + \int_0^t \sigma(\varphi(s)) \dot{g}(s) ds \right\},
 \end{aligned}$$

and define

$$I_x(\varphi) = \begin{cases} \frac{1}{2} \inf_{g \in K_\varphi} \int_0^T (\dot{g}(t))^2 dt, & \varphi \in H_1^x, \\ \infty, & \varphi \notin H_1^x. \end{cases} \quad (3.5)$$

By using the explicit expression of $D^{-1}(x)$, we get the following for functions $\varphi(t)$ whose paths are completely contained inside \mathcal{S}_n° :

$$\begin{aligned} & [\dot{\varphi}(t) - b(\varphi(t))] D^{-1}(\varphi(t)) [\dot{\varphi}(t) - b(\varphi(t))]' \\ &= \sum_{i,j=1, i \neq j}^{n-1} [\dot{\varphi}_i(t) - b_i(\varphi(t))] [\dot{\varphi}_j(t) - b_j(\varphi(t))] \left(1 - \sum_{k=1}^{n-1} \varphi_k(t) \right)^{-1} \\ & \quad + \sum_{i=1}^{n-1} [\dot{\varphi}_i(t) - b_i(\varphi(t))]^2 \left(\frac{1 - \sum_{k=1, k \neq i}^{n-1} \varphi_k(t)}{\varphi_i(t)(1 - \sum_{k=1}^{n-1} \varphi_k(t))} \right) \\ &= \sum_{i=1}^n \frac{(\dot{\varphi}_i(t) - b_i(\varphi(t)))^2}{\varphi_i(t)}, \end{aligned} \quad (3.6)$$

where $\varphi_n(t) = 1 - \sum_{i=1}^{n-1} \varphi_i(t)$.

Lemma 3.2. *If φ hits the boundary $\partial \mathcal{S}_n$, then $I_x(\varphi) = \infty$.*

Proof. We will prove this result by contradiction. Assume there is a φ such that $I_x(\varphi) < \infty$ and φ hits the boundary. Let $t_0 \in (0, T]$ be the first time that φ hits the boundary. Without loss of generality, we further assume that the hitting occurs on the first coordinate φ_1 . Now choose $0 < t_1 < t_2 < t_0$ such that $\inf_{t \in [t_1, t_2]} \{b_1(\varphi_1(t))\} > 0$, and $\log(\varphi_1(t))$ is absolutely continuous on $[t_1, t_2]$. By direct calculation we get

$$-2 \int_{t_1}^{t_2} \frac{\dot{\varphi}_1(t) b_1(\varphi(t))}{\varphi_1(t)} dt \leq \int_{t_1}^{t_2} \frac{(\dot{\varphi}_1(t) - b_1(\varphi(t)))^2}{\varphi_1(t)} dt \leq I_x(\varphi) < \infty.$$

On the other hand, performing integration by parts twice, we get

$$\begin{aligned} & -2 \int_{t_1}^{t_2} \frac{\dot{\varphi}_1(t) b_1(\varphi(t))}{\varphi_1(t)} dt \\ &= -2 \left\{ b_1(\varphi(t_2)) \log(\varphi_1(t_2)) - b_1(\varphi(t_1)) \log(\varphi_1(t_1)) \right. \\ & \quad \left. + \left\{ \frac{\theta}{2} \int_{t_1}^{t_2} \log(\varphi_1(t)) \dot{\varphi}_1(t) dt \right\} \right\} \\ &= -2 \left\{ b_1(\varphi(t_2)) \log(\varphi_1(t_2)) - b_1(\varphi(t_1)) \log(\varphi_1(t_1)) \right. \\ & \quad \left. + \left\{ \frac{\theta}{2} [\varphi_1(t_2)(\log(\varphi_1(t_2)) - 1) - \varphi_1(t_1)(\log(\varphi_1(t_1)) - 1)] \right\} \right\}. \end{aligned}$$

Letting t_2 goes to t_0 , the above will go to infinity because the first term goes to infinity while all other terms are bounded. This certainly contradicts the assumption of finiteness of $I_x(\varphi)$. \square

Hence if we define $H_0^x = H_1^x \cap C([0, T], \mathcal{S}_n^\circ)$, then

$$I_x(\varphi) = \begin{cases} \frac{1}{2} \int_0^T \sum_{i=1}^n \frac{(\dot{\varphi}_i(t) - b_i(\varphi(t)))^2}{\varphi_i(t)} dt, & \varphi \in H_0^x, \\ \infty, & \varphi \notin H_0^x. \end{cases} \quad (3.7)$$

Theorem 3.3. For any $x \in \mathcal{S}_n^\circ$, the family $\{P_x^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP on space $C([0, T], \mathcal{S}_n)$ with the good rate function $I_x(\cdot)$ and speed ε^2 .

Remark. If we introduce a map Ψ between spaces \mathcal{S}_n and Δ_n such that

$$\Psi(x_1, \dots, x_{n-1}) = \left(x_1, \dots, x_{n-1}, 1 - \sum_{i=1}^{n-1} x_i \right),$$

and for simplicity, let $\{P_x^\varepsilon\}_{\varepsilon>0}$ denote its image probability on space $C([0, T], \Delta_n)$ under the map Ψ , then by contraction principle we have that for any $x \in \Delta_n^\circ$, the family $\{P_x^\varepsilon\}_{\varepsilon>0}$ satisfies a large deviation principle on space $C([0, T], \Delta_n)$ with the good rate function $I_x(\cdot)$ and speed ε^2 .

Proof. By Corollary 3.4 in Pukhalskii (1991) and Lemma 3.1, it suffices to show that for every $\varphi \in C_x([0, T], \mathcal{S}_n)$,

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \log P_x^\varepsilon \left\{ \sup_{t \in [0, T]} |x(t) - \varphi(t)| \leq \delta \right\} \\ &= \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log P_x^\varepsilon \left\{ \sup_{t \in [0, T]} |x(t) - \varphi(t)| \leq \delta \right\} = -I_x(\varphi). \end{aligned} \quad (3.8)$$

First, we assume that the path of φ is contained in \mathcal{S}_n° . Choose δ_0 small enough such that

$$B \equiv \left\{ x(\cdot) \in C_x([0, T], \mathcal{S}_n); \sup_{t \in [0, T]} |x(t) - \varphi(t)| \leq \delta_0 \right\} \subset C_x([0, T], \mathcal{S}_n^\circ).$$

Then, we have

$$\alpha \equiv \frac{1}{2} \inf_{t \in [0, T], x(\cdot) \in B} d(x(t), \partial \mathcal{S}_n) > 0,$$

where $d(y, \partial \mathcal{S}_n) = \inf_{z \in \partial \mathcal{S}_n} |y - z|$.

Define

$$\tilde{\sigma}(x) \equiv \begin{cases} \sigma(x), & d(x, \partial \mathcal{S}_n) \geq \alpha \\ \text{smooth and uniform Lipschitz continuous,} & \text{else.} \end{cases}$$

By replacing $\sigma(x)$ with $\tilde{\sigma}(x)$, we define the diffusion $\tilde{x}^\varepsilon(t)$, the law \tilde{P}_x^ε , the matrix $\tilde{A}(x)$, and the function $\tilde{I}_x(\cdot)$, respectively. It is easy to see that $I_x(\varphi) = \tilde{I}_x(\varphi)$. By using Theorem 5.6.7 in Dembo and Zeitouni (1993) and Theorem 3.5 of Ch. 3 in Freidlin and Wentzell (1984), we have for any $f \in C_x([0, T], \mathcal{S}_n)$,

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \log \tilde{P}_x^\varepsilon \left\{ \sup_{t \in [0, T]} |x(t) - f(t)| \leq \delta \right\} \\ &= \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log \tilde{P}_x^\varepsilon \left\{ \sup_{t \in [0, T]} |x(t) - f(t)| \leq \delta \right\} = -\tilde{I}_x(f). \end{aligned} \quad (3.9)$$

Replacing f with φ in Eq. (3.9), we end up with Eq. (3.8) because the corresponding probabilities in both equations are the same for $\delta < \delta_0$.

Now, we are ready to prove Eq. (3.8) for paths that hit the boundary of \mathcal{S}_n . Let φ be such a path. The following is trivially true.

$$\lim_{\delta \rightarrow 0} \liminf_{\varepsilon \rightarrow 0} \varepsilon^2 \log P_x^\varepsilon \left\{ \sup_{t \in [0, T]} |x(t) - \varphi(t)| \leq \delta \right\} \geq -I_x(\varphi) = -\infty. \quad (3.10)$$

On the other hand, let $t_0 > 0$ be the first time when φ hits the boundary. Then for any $t \in (0, t_0)$, φ will not hit the boundary on $[0, t]$. By using an argument similar to that used in the derivation of Eq. (3.9) we get

$$\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log \tilde{P}_x^\varepsilon \left\{ \sup_{s \in [0, t]} |x(s) - \varphi(s)| \leq \delta \right\} = -I_x^t(\varphi), \quad (3.11)$$

where $I_x^t(\varphi)$ is the restriction of $I_x(\varphi)$ on $[0, t]$. Hence,

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log \tilde{P}_x^\varepsilon \left\{ \sup_{s \in [0, T]} |x(s) - \varphi(s)| \leq \delta \right\} \\ & \leq \lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log \tilde{P}_x^\varepsilon \left\{ \sup_{s \in [0, t]} |x(s) - \varphi(s)| \leq \delta \right\} = -I_x^t(\varphi). \end{aligned} \quad (3.12)$$

From the proof of Lemma 3.2, we get that $I_x^t(\varphi)$ converges to infinity as $t \nearrow t_0$ which implies

$$\lim_{\delta \rightarrow 0} \limsup_{\varepsilon \rightarrow 0} \varepsilon^2 \log \tilde{P}_x^\varepsilon \left\{ \sup_{s \in [0, T]} |x(s) - \varphi(s)| \leq \delta \right\} = -\infty. \quad (3.13)$$

Lemma 3.2 combined with Eqs. (3.10) and (3.13) implies that Eq. (3.8) holds for all $\varphi \in C([0, T], \mathcal{S}_n)$. \square

3.2. LDP for Fleming–Viot processes with neutral mutation

The Fleming–Viot process with neutral mutation has many nice properties. One of them is called the partition property, namely, given any finite partition of the type space $E = \bigcup_{i=1}^K E_i$, then $\{X(t, E_i) : i = 1, \dots, K\}$ is a finite-dimensional diffusion process as

in Section 3.1 (cf. Ethier and Kurtz, 1994.) By using this property and the projective limit technique, we will establish a LDP for the Fleming–Viot process with neutral mutation. This kind of result can be viewed as an infinite dimensional generalization of the Freidlin–Wentzell theory. In the remainder of this article we will have $E = [0, 1]$.

Consider the following family of partitions of space E

$$\mathcal{J} = \{ \{[0, t_1], (t_1, t_2], \dots, (t_n, 1]\} : 0 < t_1 < t_2 < \dots < t_n < 1; n = 1, 2, \dots \},$$

with the same partial ordering \succ as in \mathcal{P} of Eq. (2.12). It is clear that $\mathcal{J} \subset \mathcal{P}$. Thus, we will denote generic element of \mathcal{J} by ι, j , etc., and for any $\iota, j \in \mathcal{J}$, the space \mathcal{X}_ι and the mappings $\pi_{\iota j}, \pi_\iota$ are defined accordingly.

By an argument similar to that used in Section 2, we can show that the projective limit of $(\mathcal{X}_\iota, \pi_{\iota j})_{\iota, j \in \mathcal{J}}$ is $M_1(E)$ and projective topology is stronger than the weak topology but weaker than the τ -topology. We will use $M_1^{\text{pro}}(E)$ to denote the space $M_1(E)$ with this projective limit topology.

Let $\mathcal{C}_j = C([0, T], \mathcal{X}_j)$ be equipped with the usual uniform topology. For any $j = \{B_1, \dots, B_r\} \succ \iota = \{C_1, \dots, C_l\}$, define a map $p_{\iota j}$ between spaces \mathcal{C}_j and \mathcal{C}_ι such that

$$p_{\iota j} : \mathcal{C}_j \rightarrow \mathcal{C}_\iota, \quad (x_{B_1}(t), \dots, x_{B_r}(t)) \rightarrow \left(\sum_{B_k \subset C_1} x_{B_k}(t), \dots, \sum_{B_k \subset C_l} x_{B_k}(t) \right).$$

It is clear that $p_{\iota j}$ is continuous, and for any $\ell \succ j \succ \iota$, $p_{\iota \ell} = p_{\iota j} \circ p_{j \ell}$. Now, let \mathcal{C} be the projective limit of the family $\{(\mathcal{C}_\iota, p_{\iota j}); \iota, j \in \mathcal{J}\}$ and $p_j : \mathcal{C} \rightarrow \mathcal{C}_j$ be the corresponding projection. Obviously, $C([0, T], M_1^{\text{pro}}([0, 1]))$ is a subset of \mathcal{C} . On the other hand, let $\{\mu^j(\cdot) : j \in \mathcal{J}\}$ be any element of \mathcal{C} . Then for any $t \in [0, T]$, $\{\mu^j(t) : j \in \mathcal{J}\}$ can be identified as a unique element $\mu(t)$ of $M_1^{\text{pro}}(E)$ and thus $\{\mu^j(\cdot) : j \in \mathcal{J}\}$ can be identified as $\mu(\cdot)$. For any $(a, b) \subset E$, by the definition of projective limit, $\mu(t)((a, b))$ is continuous in t . Hence $\mu(\cdot) \in C([0, T], M_1^{\text{pro}}(E))$ and \mathcal{C} can be identified as $C([0, T], M_1^{\text{pro}}(E))$.

For any $v \in M_1(E)$, let $P_v^{\theta, \gamma, v_0}$ be the unique solution of the martingale problem associated with generator \mathcal{L} in Eq. 2.15 starting at v .

Definition 3.1. A probability measure $v \in M_1(E)$ is called non-degenerate if for any $j \in \mathcal{J}$, $\pi_j(v)$ has no zero component. A path $\mu(\cdot) \in C([0, T], M_1^{\text{pro}}(E))$ is called non-degenerate if for every t in $[0, T]$, $\mu(t)$ is non-degenerate.

Remark. Any probability measure with support E is non-degenerate.

Definition 3.2. A path $\mu(\cdot) \in C([0, T], M_1^{\text{pro}}(E))$ is called absolutely continuous if for every $\iota \in \mathcal{J}$, $\pi_\iota(\mu)(t)$ is absolutely continuous as a multidimensional real-valued function.

Theorem 3.4. For any non-degenerate $v \in M_1(E)$, the family $\{P_v^{\theta, \gamma, v_0}\}$ satisfies a LDP on space $C([0, T], M_1^{\text{pro}}(E))$ as γ goes to zero with the good rate function

$$I_v(\mu(\cdot)) = \sup_{j \in \mathcal{J}} I_{\pi_j(v)}(p_j(\mu(\cdot))). \quad (3.14)$$

Remark. Let $\mu(t)$ be a path of the Fleming–Viot process with neutral mutation. Then for every $t > 0$, the support of $\mu(t)$ is a subset of the support of ν_0 . Therefore, the essential part of the type space is the support of ν_0 . If we choose E to be the support of ν_0 , then the result still holds. Because of this we assume in the sequel, without loss of generality, that the support of ν_0 is $E = [0, 1]$.

Proof. Since the path of Fleming–Viot process is continuous in the τ -topology (cf. Shiga, 1990), we have that $P_v^{\theta, \gamma, \nu_0} \{C([0, T], M_1^{\text{pro}}(E))\} = 1$. On the other hand, for any $j \in \mathcal{J}$, by the nondegeneracy of ν , Theorem 3.3 and the remark following it, we have that $p_j(P_v^{\theta, \gamma, \nu_0})$ satisfies a LDP on space \mathcal{C}_j with the good rate function $I_{\pi_j(\nu)}(\cdot)$. Applying Theorem 3.3 in Dawson and Gärtner (1987) we get the result. \square

Let $C^{1,0}([0, T] \times E)$ denote the set of all continuous functions on $[0, T] \times E$ with continuous first order derivative in time. For any $\nu \in M_1(E)$, $\mu(\cdot) \in C([0, T], M_1^{\text{pro}}(E))$, define

$$\begin{aligned} S_\nu(\mu(\cdot)) = \sup_{f \in C^{1,0}([0, T] \times E)} & \left\{ \langle \mu(T), f(T) \rangle - \langle \mu(0), f(0) \rangle \right. \\ & - \int_0^T \left\langle \mu(s), \left(\frac{\partial}{\partial s} + A \right) f \right\rangle ds \\ & \left. - \frac{1}{2} \int_0^T \int \int f(s, x) f(s, y) Q(\mu(s); dx, dy) ds \right\}, \end{aligned} \quad (3.15)$$

and

$$\begin{aligned} \mathcal{H}_\nu = \{ \mu(\cdot) \in C([0, T], M_1^{\text{pro}}(E)) : \mu(0) = \nu, \mu(\cdot) \text{ is absolutely continuous,} \\ \text{non-degenerate, and for any } j \in \mathcal{J}, p_j(\mu)(\cdot) \in L^2([0, T]) \}. \end{aligned}$$

Remark. It is clear that $I_\nu(\mu(\cdot)) = \infty$ for any $\mu(\cdot)$ not in \mathcal{H}_ν .

The next theorem gives an variational form of the rate function $I_\nu(\mu(\cdot))$.

Theorem 3.5. For any non-degenerate $\nu \in M_1(E)$, and any $\mu(\cdot) \in \mathcal{H}_\nu$, we have

$$I_\nu(\mu(\cdot)) = S_\nu(\mu(\cdot)).$$

Proof. For any

$$j = \{B_1 = [0, b_1), B_2 = [b_1, b_2), \dots, B_{n-1} = [b_{n-1}, 1]\} \in \mathcal{J},$$

and $f(t, y) \in C^{1,0}([0, T] \times E)$, let $b_0 = 0$ and

$$\pi_j(f)(t, y) = f(t, b_k), \quad \text{for } y \in B_k, \quad k = 0, \dots, n-1.$$

Then from the absolute continuity of $\mu(\cdot)$, we have

$$\begin{aligned} \int_0^T \langle \dot{\mu}(s), \pi_j(f)(s) \rangle ds &= \langle \mu(T), \pi_j(f)(T) \rangle - \langle \mu(0), \pi_j(f)(0) \rangle \\ &\quad - \int_0^T \left\langle \mu(s), \frac{\partial}{\partial s} \pi_j(f)(s) \right\rangle ds. \end{aligned} \quad (3.16)$$

Approximating f by $\pi_j(f)$, we obtain the following result of integration by parts:

$$\int_0^T \langle \dot{\mu}(s), f(s) \rangle ds = \langle \mu(T), f(T) \rangle - \langle \mu(0), f(0) \rangle - \int_0^T \langle \mu(s), \dot{f}(s) \rangle ds. \quad (3.17)$$

Let $x = \pi_j(v)$. For $k=0, \dots, n-1$, let $\varphi_k(t) = p_j(\mu)(t)(B_k)$, $\varphi(t) = (\varphi_0(t), \dots, \varphi_{n-1}(t))$ and $g(t, y) = (\dot{\varphi}_k(t) - b_k(\varphi(t)))^2 / \varphi_k(t)$ for $y \in B_k$. Here $g(t, y)$ is well defined since $\mu(\cdot)$ is non-degenerate, and absolutely continuous. Then by direct calculation we have

$$\begin{aligned} I_{\pi_j(v)}(p_j(\mu(\cdot))) &= I_x(\varphi) = \frac{1}{2} \sum_{k=0}^{n-1} \int_0^T \frac{(\dot{\varphi}_k(s) - b_k(\varphi(s)))^2}{\varphi_k(s)} ds \\ &= \int_0^T \langle \dot{\mu}(s), g(s) \rangle ds - \int_0^T \langle \mu(s), Ag \rangle ds \\ &\quad - \frac{1}{2} \int_0^T \int_E \int_E g(s, x) g(s, y) Q(\mu(s); dx, dy) ds \\ &\leq \sup_{f \in C^{1,0}([0, T] \times E)} \left\{ \int_0^T \langle \dot{\mu}(s), \pi_j(f)(s) \rangle ds - r \int_0^T \langle \mu(s), A(\pi_j(f)) \rangle ds \right. \\ &\quad \left. - \frac{1}{2} \int_0^T \int_E \int_E \pi_j(f)(s, x) \pi_j(f)(s, y) Q(\mu(s); dx, dy) ds \right\}, \end{aligned} \quad (3.18)$$

where the inequality holds because g can be approximated pointwise by functions in the set $\{\pi_j(f): f \in C^{1,0}([0, T] \times E)\}$.

On the other hand,

$$\begin{aligned} &\sup_{f \in C^{1,0}([0, T] \times E)} \left\{ \int_0^T \langle \dot{\mu}(s), \pi_j(f)(s) \rangle ds - \int_0^T \langle \mu(s), A(\pi_j(f)) \rangle ds \right. \\ &\quad \left. - \frac{1}{2} \int_0^T \int_E \int_E \pi_j(f)(s, x) \pi_j(f)(s, y) Q(\mu(s); dx, dy) ds \right\} \\ &\leq \sup_{f \in C^{1,0}([0, T] \times E)} \int_0^T \left[\sum_{k=1}^{n-1} f(s, B_k) (\dot{\varphi}_k(s) - b_k(\varphi(s))) \right. \\ &\quad \left. - \frac{1}{2} \sum_{k, l=1}^{n-1} f(s, B_k) D_{kl}(\varphi(s)) f(s, B_l) \right] ds \\ &\leq \int_0^T \sup_{\alpha \in R^n} \left[\sum_{k=1}^{n-1} \alpha_k (\dot{\varphi}_k(s) - b_k(\varphi(s))) - \frac{1}{2} \sum_{k, l=1}^{n-1} \alpha_k D_{kl}(\varphi(s)) \alpha_l \right] ds \\ &= I_x(\varphi), \end{aligned} \quad (3.19)$$

which, together with Eq. (3.18), implies that

$$I_{\pi_j(v)}(p_j(\mu(\cdot))) = \sup_{f \in C^{1,0}([0,T]) \times E} \left\{ \int_0^T \langle \dot{\mu}(s), \pi_j(f)(s) \rangle \, ds - \int_0^T \langle \mu(s), A(\pi_j(f)) \rangle \, ds \right. \\ \left. - \frac{1}{2} \int_0^T \int_E \int_E \pi_j(f)(s,x) \pi_j(f)(s,y) Q(\mu(s); dx, dy) \, ds \right\}. \quad (3.20)$$

By using the inequality

$$\frac{(a+b)^2}{c+d} \leq \frac{a^2}{c} + \frac{b^2}{d}, \quad a, b, c, d > 0,$$

and expression (3.7) for $I_x(\varphi)$, we have that for any $j_2 \succ j_1$,

$$I_{\pi_{j_2}(v)}(p_{j_2}(\mu(\cdot))) \geq I_{\pi_{j_1}(v)}(p_{j_1}(\mu(\cdot))).$$

Taking the supremum on both sides of Eqs. (3.20) over the set \mathcal{J} , we finally get

$$I_v(\mu(\cdot)) = \sup_{f \in C^{1,0}([0,T]) \times E} \left\{ \int_0^T \langle \dot{\mu}(s), f(s) \rangle \, ds - \int_0^T \langle \mu(s), Af \rangle \, ds \right. \\ \left. - \frac{1}{2} \int_0^T \int_E \int_E f(s,x) f(s,y) Q(\mu(s); dx, dy) \, ds \right\} \\ = S_v(\mu(\cdot)), \quad (3.21)$$

where the last equality follows from integration by parts (3.17). \square

Let $C^\infty(E)$ be the family of all continuous functions on E possessing continuous derivatives of all order. For any linear functional ϑ on space $C^\infty(E)$, define

$$\|\vartheta\|_\mu^2 = \sup_{f \in C^\infty(E)} \left[\langle \vartheta, f \rangle - \frac{1}{2} \int_E \int_E f(x) f(y) Q(\mu; dx, dy) \right].$$

Then we have

Theorem 3.6. *For any non-degenerate $v \in M_1(E)$, and any $\mu(\cdot)$ in \mathcal{H}_v , we have*

$$I_v(\mu(\cdot)) = \int_0^T \|\dot{\mu}(s) - A^*(\mu(s))\|_{\mu(s)}^2 \, ds, \quad (3.22)$$

where A^* is the formal adjoint of A define through the equality $\langle A^*(\mu), f \rangle = \langle \mu, Af \rangle$.

We defer the proof of this theorem to the appendix.

3.3. LDP for Fleming–Viot processes with selection

Finally, we turn to prove the LDP of the Fleming–Viot process with selection. The generator of the process is given in Eq. (2.16). We will assume that the fitness function $V(x, y)$ is continuous on $E^{\otimes 2}$ in the sequel.

For any $v \in M_1(E)$, let $P_v^{\theta, \gamma, V, v_0}$ be the unique solution of the martingale problem associated with generator \mathcal{L}_V started at v . For simplicity, we will not distinguish between $P_v^{\theta, \gamma, v_0}$, $P_v^{\theta, \gamma, V, v_0}$ and their respective restrictions on $C([0, T], M_1^{\text{pro}}(E))$. By the Cameron–Martin–Girsanov transformation (see Dawson, 1978) we have that, restricted on $C([0, T], M_1^{\text{pro}}(E))$,

$$\frac{dP_v^{\theta, \gamma, v_0}}{dP_v^{\theta, \gamma, V, v_0}} = Z_V(T) = \exp \left[\frac{1}{\gamma} G_V(\mu(\cdot)) \right] > 0, \quad (3.23)$$

where

$$\begin{aligned} G_V(\mu(\cdot)) = & \int_0^T \int_E \left[\int_E V(y, z) \mu(s, dz) \right] M(ds, dy) \\ & - \frac{1}{2} \int_0^T \int_E \int_E \left[\int_E V(x, z) \mu(s, dz) \right] \left[\int_E V(y, z) \mu(s, dz) \right] \\ & \times Q(\mu(s); dx, dy) ds, \end{aligned} \quad (3.24)$$

and $M(ds, dy)$ is the martingale measure obtained from the martingale

$$M_t(\phi) = \langle \mu(t), \phi \rangle - \langle \mu(0), \phi \rangle - \int_0^t \langle \mu(s), A\phi \rangle ds.$$

Let ρ be the Prohorov metric on $M_1(E)$, and $C^w([0, T], M_1^{\text{pro}}(E))$ denote $C([0, T], M_1^{\text{pro}}(E))$ equipped with the subspace topology of $C([0, T], M_1(E))$. For any $v(\cdot), \mu(\cdot) \in C^w([0, T], M_1^{\text{pro}}(E))$, let

$$\varrho(v(\cdot), \mu(\cdot)) = \sup_{t \in [0, T]} \rho(v(t), \mu(t)).$$

Define

$$R(\mu, dx) = \int_E \left[\int_E V(y, z) \mu(dz) \right] Q(\mu; dx, dy).$$

Then we have the following theorem.

Theorem 3.7. *For any non-degenerate $v \in M_1(E)$, the family $\{P_v^{\theta, \gamma, V, v_0}\}$ satisfies the following local LDP on space $C^w([0, T], M_1^{\text{pro}}(E))$ as γ goes to zero with the good rate function $I_{v, V}(\mu(\cdot))$:*

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \liminf_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, V, v_0} \{ \varrho(v(\cdot), \mu(\cdot)) < \delta \} \\ & = \lim_{\delta \rightarrow 0} \limsup_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, V, v_0} \{ \varrho(v(\cdot), \mu(\cdot)) \leq \delta \} = -I_{v, V}(\mu(\cdot)), \end{aligned} \quad (3.25)$$

where

$$I_{v, V}(\mu(\cdot)) = \begin{cases} \int_0^T \|\dot{\mu}(s) - R(\mu(s)) - A^*(\mu(s))\|_{\mu(s)}^2 ds, & \text{if } \mu(0) = v \text{ and } \mu(\cdot) \text{ is} \\ & \text{absolutely continuous,} \\ \infty, & \text{elsewhere.} \end{cases} \quad (3.26)$$

Remark. This yields the weak LDP, that is, the upper bound holds for compact sets only.

Proof. Theorem 3.4 combined with Theorem 3.5 in Freidlin and Wentzell (1984), we have for $\mu(\cdot) \in C([0, T], M_1^{\text{pro}}(E))$,

$$\begin{aligned} & \lim_{\delta \rightarrow 0} \liminf_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{v: \varrho(v(\cdot), \mu(\cdot)) < \delta\} \\ &= \lim_{\delta \rightarrow 0} \limsup_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{v: \varrho(v(\cdot), \mu(\cdot)) \leq \delta\} = -I_v(\mu(\cdot)). \end{aligned} \quad (3.27)$$

Let

$$\begin{aligned} C = & \sup_{\mu(\cdot) \in C^w([0, T], M_1^{\text{pro}}(E))} \left| \int_0^T \int_E \int_E \left[\int_E V(x, z) \mu(s, dz) \right] \right. \\ & \times \left. \left[\int_E V(y, z) \mu(s, dz) \right] \mathcal{Q}(\mu(s); dx, dy) ds \right|. \end{aligned}$$

Since V is uniformly bounded, C is a finite constant. For any measurable subset B of space $C^w([0, T], M_1^{\text{pro}}(E))$, by using Eq. (3.23), Hölder's inequality, and martingale property, we get for any $\alpha > 0, \beta > 0, 1/\alpha + 1/\beta = 1$,

$$\begin{aligned} P_v^{\theta, \gamma, V, v_0} \{B\} &= \int_B Z_V(T) dP_v^{\theta, \gamma, v_0} \\ &\leq e^{C/2\gamma} \left(\int \exp \left[\frac{\alpha}{\gamma} \int_0^T \int_E \left(\int_E V(y, z) \mu(s, dz) \right) M(ds, dy) \right] dP_v^{\theta, \gamma, v_0} \right)^{1/\alpha} \\ &\quad \times P_v^{\theta, \gamma, v_0} \{B\}^{1/\beta} \\ &\leq e^{(C/2\gamma)(1+\alpha)} P_v^{\theta, \gamma, v_0} \{B\}^{1/\beta} \left(\int Z_{\alpha V}(T) dP_v^{\theta, \gamma, v_0} \right)^{1/\alpha} \\ &= e^{(C/2\gamma)(1+\alpha)} P_v^{\theta, \gamma, v_0} \{B\}^{1/\beta}. \end{aligned} \quad (3.28)$$

By choosing $B = \{v: \varrho(v(\cdot), \mu(\cdot)) \leq \delta\}$ in Eq. (3.28), we get

$$\lim_{\delta \rightarrow 0} \limsup_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, V, v_0} \{\varrho(v(\cdot), \mu(\cdot)) \leq \delta\} \leq \frac{C(1+\alpha)}{2} - \frac{1}{\beta} I_v(\mu(\cdot)) \quad (3.29)$$

which implies Eq. (3.25) for paths that are not absolutely continuous.

Let $\tilde{\mu}(\cdot)$ be an arbitrary absolutely continuous path in $C^w([0, T], M_1^{\text{pro}}(E))$. For any $\mu \in M_1(E)$, let $V(\mu)(y) = V(\mu, y) = \int_E V(y, z) \mu(dz)$. Then we have

$$\begin{aligned} \int_0^T \int_E V(\tilde{\mu}(s), y) M(ds, dy) &= \langle \mu(T), V(\tilde{\mu}(T)) \rangle - \langle \mu(0), V(\tilde{\mu}(0)) \rangle \\ &\quad - \int_0^T \left\langle \mu(s), \left(\frac{\partial}{\partial s} + A \right) V(\tilde{\mu}(s)) \right\rangle ds, \end{aligned} \quad (3.30)$$

which is continuous in $\mu(\cdot)$ in the topology generated by ϱ . On the other hand,

$$\begin{aligned} \int_E \int_E V(\mu, x) V(\mu, y) Q(\mu; dx, dy) &= \int_E \int_E \int_E V(x, z_1) V(x, z_2) \mu(dz_1) \mu(dz_2) \mu(dx) \\ &\quad - \int_E \int_E \int_E \int_E V(x, z_1) V(y, z_2) \\ &\quad \times \mu(dz_1) \mu(dz_2) \mu(dx) \mu(dy), \end{aligned}$$

which is also continuous in the topology generated by ϱ .

Define

$$\begin{aligned} \Gamma_V(\mu(\cdot)) &= \langle \mu(T), V(\mu(T)) \rangle - \langle \mu(0), V(\mu(0)) \rangle \\ &\quad - \int_0^T \left\langle \mu(s), \left(\frac{\partial}{\partial s} + A \right) V(\mu(s)) \right\rangle ds \\ &\quad - \frac{1}{2} \int_0^T \int_E \int_E \left[\int_E V(x, z) \mu(s, dz) \right] \left[\int_E V(y, z) \mu(s, dz) \right] \\ &\quad \times Q(\mu(s); dx, dy) ds. \end{aligned} \quad (3.31)$$

Let $B^\circ(\bar{\mu}(\cdot), \delta)$ denote the interior of $B(\bar{\mu}(\cdot), \delta)$. For any $\varepsilon > 0$, we can choose $\delta > 0$ small enough such that

$$\left| \liminf_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{B^\circ(\bar{\mu}(\cdot), \delta)\} + I_v(\bar{\mu}(\cdot)) \right| < \varepsilon, \quad (3.32)$$

$$\left| \limsup_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{B(\bar{\mu}(\cdot), \delta)\} + I_v(\bar{\mu}(\cdot)) \right| < \varepsilon, \quad (3.33)$$

$$\begin{aligned} \left| \int_0^T \int_E V(\bar{\mu}(s), y) M(ds, dy) - \left(\langle \bar{\mu}(T), V(\bar{\mu}(T)) \rangle - \langle \bar{\mu}(0), V(\bar{\mu}(0)) \rangle \right) \right. \\ \left. - \int_0^T \left\langle \bar{\mu}(s), \left(\frac{\partial}{\partial s} + A \right) V(\bar{\mu}(s)) \right\rangle ds \right| < \varepsilon, \end{aligned} \quad (3.34)$$

$$\begin{aligned} \left| \int_0^T \int_E \int_E V(\mu(s), x) V(\mu(s), y) Q(\mu(s); dx, dy) ds \right. \\ \left. - \int_0^T \int_E \int_E V(\bar{\mu}(s), x) V(\bar{\mu}(s), y) Q(\bar{\mu}(s); dx, dy) ds \right| < \varepsilon, \end{aligned} \quad (3.35)$$

$$\begin{aligned} \left| \int_0^T \int_E \int_E V(\bar{\mu}(s), x) V(\bar{\mu}(s), y) Q(\mu; dx, dy) ds \right. \\ \left. - \int_0^T \int_E \int_E V(\bar{\mu}(s), x) V(\bar{\mu}(s), y) Q(\bar{\mu}(s); dx, dy) ds \right| < \varepsilon, \end{aligned} \quad (3.36)$$

$$\left| \int_0^T \int_E \int_E (V(\mu(s), x) - V(\bar{\mu}(s), x))(V(\mu(s), y) - V(\bar{\mu}(s), y)) Q(\mu(s); dx, dy) ds \right| < \varepsilon. \quad (3.37)$$

By Eqs. (3.34)–(3.37), we get

$$\begin{aligned}
 & P_v^{\theta, \gamma, V, v_0} \{B(\bar{\mu}(\cdot), \delta)\} \\
 &= \int_{B(\bar{\mu}(\cdot), \delta)} \exp \left[\frac{1}{\gamma} G_V(\mu(\cdot)) \right] dP_v^{\theta, \gamma, v_0} \\
 &\leq \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) + \frac{3}{2} \varepsilon \right] \right\} \\
 &\quad \times \int_{B(\bar{\mu}(\cdot), \delta)} \exp \left\{ \frac{1}{\gamma} \left[\int_0^T \int_E \left[\int_E V(y, z)(\mu(ds, dz) \right. \right. \right. \\
 &\quad \left. \left. \left. - \bar{\mu}(s, dz) \right) \right] M(ds, dy) \right] \right\} dP_v^{\theta, \gamma, v_0} \\
 &\leq \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) + \frac{3 + \alpha}{2} \varepsilon \right] \right\} P_v^{\theta, \gamma, v_0} \{B(\bar{\mu}(\cdot), \delta)\}^{1/\beta} \\
 &\quad \times \left(\int \exp \left\{ \frac{1}{\gamma} \left[\int_0^T \int_E \left[\int_E \alpha V(y, z)(\mu(ds, dz) - \bar{\mu}(s, dz)) \right] M(ds, dy) \right. \right. \right. \\
 &\quad \left. \left. \left. - \frac{\alpha^2}{2} \int_0^T \int_E \int_E (V(\mu(s), x) - V(\bar{\mu}(s), x)) \right. \right. \right. \\
 &\quad \left. \left. \left. \times \left[\left\{ \left((V(\mu(s), y) - V(\bar{\mu}(s), y)) Q(\mu(s); dx, dy) ds \right) \right\} dP_v^{\theta, \gamma, v_0} \right] \right\} \right)^{1/\alpha} \\
 &= \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) + \frac{3 + \alpha}{2} \varepsilon \right] \right\} P_v^{\theta, \gamma, v_0} \{B(\bar{\mu}(\cdot), \delta)\}^{1/\beta}. \tag{3.38}
 \end{aligned}$$

Together with Eq. (3.33) this implies

$$\lim_{\delta \rightarrow 0} \limsup_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{B(\bar{\mu}(\cdot), \delta)\} \leq \Gamma_V(\bar{\mu}(\cdot)) + \frac{3 + \alpha}{2} \varepsilon - \frac{1}{\beta} I_V(\bar{\mu}(\cdot)).$$

By letting $\varepsilon \rightarrow 0$, then $\beta \rightarrow 1$, we get

$$\lim_{\delta \rightarrow 0} \limsup_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{B(\bar{\mu}(\cdot), \delta)\} \leq \Gamma_V(\bar{\mu}(\cdot)) - I_V(\bar{\mu}(\cdot)). \tag{3.39}$$

By an argument similar to that used in the derivation of Eq. (3.38) and Hölder's inequality we get

$$\begin{aligned}
 & P_v^{\theta, \gamma, V, v_0} \{B^\circ(\bar{\mu}(\cdot), \delta)\} \\
 &= \int_{B^\circ(\bar{\mu}(\cdot), \delta)} \exp \left[\frac{1}{\gamma} G_V(\mu(\cdot)) \right] dP_v^{\theta, \gamma, v_0} \\
 &\geq \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) - \frac{3}{2} \varepsilon \right] \right\} \\
 &\quad \times \int_{B^\circ(\bar{\mu}(\cdot), \delta)} \exp \left\{ \frac{1}{\gamma} \left[\int_0^T \int_E \left[\int_E V(y, z)(\mu(ds, dz) \right. \right. \right. \\
 &\quad \left. \left. \left. - \bar{\mu}(s, dz) \right) \right] M(ds, dy) \right] \right\} dP_v^{\theta, \gamma, v_0}
 \end{aligned}$$

$$\begin{aligned}
&\geq \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) - \left(\frac{3}{2} + \frac{\alpha}{2\beta} \right) \varepsilon \right] \right\} \\
&\quad \times \int_{B^\circ(\bar{\mu}(\cdot), \delta)} \exp \left\{ \frac{1}{\gamma} \left[\int_0^T \int_E \left(\int_E V(y, z)(\mu(ds, dz) \right. \right. \right. \\
&\quad \left. \left. \left. - \bar{\mu}(s, dz) \right) M(ds, dy) \right. \right. \\
&\quad \left. \left. + \frac{\alpha}{2\beta} \int_0^T \int_E \int_E (V(\mu(s), x) - V(\bar{\mu}(s), x)) \right. \right. \\
&\quad \left. \left. \times \left[\left\{ (V(\mu(s), y) - V(\bar{\mu}(s), y)) Q(\mu(s); dx, dy) ds \right\} dP_v^{\theta, \gamma, v_0} \right] \right\} \right. \\
&\geq \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) - \left(\frac{3}{2} + \frac{\alpha}{2\beta} \right) \varepsilon \right] \right\} P_v^{\theta, \gamma, v_0} \{B^\circ(\bar{\mu}(\cdot), \delta)\}^\beta \\
&\quad \times \left(\int \exp \left\{ \frac{1}{\gamma} \left[\int_0^T \int_E \left[\int_E -\frac{\alpha}{\beta} V(y, z)(\mu(ds, dz) \right. \right. \right. \right. \\
&\quad \left. \left. \left. - \bar{\mu}(s, dz) \right) M(ds, dy) \right. \right. \\
&\quad \left. \left. - \frac{\alpha^2}{2\beta^2} \int_0^T \int_E \int_E (V(\mu(s), x) - V(\bar{\mu}(s), x)) \right. \right. \\
&\quad \left. \left. \times \left(\left[\left\{ (V(\mu(s), y) - V(\bar{\mu}(s), y)) Q(\mu(s); dx, dy) ds \right\} dP_v^{\theta, \gamma, v_0} \right] \right)^{-\beta/\alpha} \right) \\
&= \exp \left\{ \frac{1}{\gamma} \left[\Gamma_V(\bar{\mu}(\cdot)) - \left(\frac{3}{2} + \frac{\alpha}{2\beta} \right) \varepsilon \right] \right\} P_v^{\theta, \gamma, v_0} \{B^\circ(\bar{\mu}(\cdot), \delta)\}^\beta, \tag{3.40}
\end{aligned}$$

which, combined with Eq. (3.32), leads to

$$\lim_{\delta \rightarrow 0} \liminf_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{B^\circ(\bar{\mu}(\cdot), \delta)\} \geq \Gamma_V(\bar{\mu}(\cdot)) - \left(\frac{3}{2} + \frac{\alpha}{2\beta} \right) \varepsilon - \beta I_v(\bar{\mu}(\cdot)).$$

By letting $\varepsilon \rightarrow 0$, then $\beta \rightarrow 1$, we get

$$\lim_{\delta \rightarrow 0} \liminf_{\gamma \rightarrow 0} \gamma \log P_v^{\theta, \gamma, v_0} \{B^\circ(\bar{\mu}(\cdot), \delta)\} \geq \Gamma_V(\bar{\mu}(\cdot)) - I_v(\bar{\mu}(\cdot)). \tag{3.41}$$

The equalities in Eq. (3.25) will follow if we can show that for absolutely continuous path $\mu(\cdot)$,

$$I_{v, V}(\mu(\cdot)) = -\Gamma_V(\mu(\cdot)) + I_v(\mu(\cdot)). \tag{3.42}$$

By direct calculation we have

$$\begin{aligned}
&I_v(\mu(\cdot)) - \Gamma_V(\mu(\cdot)) \\
&= \sup_{f \in C^{1,0}([0, T] \times E)} \left\{ \langle \mu(T), f(T) - V(\mu(T)) \rangle - \langle \mu(0), f(0) - V(\mu(0)) \rangle \right. \\
&\quad \left. - \int_0^T \left[\left\langle \mu(s), \left(\frac{\partial}{\partial s} + A \right) f - V(\mu(s)) \right\rangle + \langle R(\mu(s)), f(s) - V(\mu(s)) \rangle \right] ds \right. \\
&\quad \left. - \frac{1}{2} \int_0^T \int_E \int_E (f(s, x) - V(\mu(s), x))(f(s, y) \right.
\end{aligned}$$

$$\begin{aligned}
 & - \left\{ V(\mu(s), y)) \mathcal{Q}(\mu(s); \mathrm{d}x, \mathrm{d}y) \, \mathrm{d}s \right\} \\
 = & \sup_{f \in C^{1,0}([0, T] \times E)} \left\{ \langle \mu(T), f(T) \rangle - \langle \mu(0), f(0) \rangle \right. \\
 & - \int_0^T \left[\left\langle \mu(s), \left(\frac{\partial}{\partial s} + A \right) f \right\rangle + \langle R(\mu(s)), f(s) \rangle \right] \mathrm{d}s \\
 & \left. - \left\{ \frac{1}{2} \int_0^T \int_E \int_E f(s, x) f(s, y) \mathcal{Q}(\mu(s); \mathrm{d}x, \mathrm{d}y) \, \mathrm{d}s \right\} \right\}, \tag{3.43}
 \end{aligned}$$

which leads to Eq. (3.42).

Finally by Eq. (3.42), we get that the level sets associated with $I_{\nu, \nu}(\mu(\cdot))$ are compact. \square

Acknowledgements

We thank an anonymous referee for comments and suggestions that helped improve the presentation of this paper.

Appendix

The proof of Theorem 3.6:

For any f in $C^{1,0}([0, T] \times E)$, and $\mu(\cdot)$ in \mathcal{H}_ν , we have that

$$\begin{aligned}
 & \langle \mu(T), f(T) \rangle - \langle \mu(0), f(0) \rangle - \int_0^T \left\langle \mu(u), \left(\frac{\partial}{\partial u} + A \right) f \right\rangle \mathrm{d}u \\
 & - \frac{1}{2} \int_0^T \int_E \int_E f(u, x) f(u, y) \mathcal{Q}(\mu(u); \mathrm{d}x, \mathrm{d}y) \mathrm{d}u \\
 = & \int_0^T \langle \dot{\mu}(u) - A^*(\mu(u)), f(u) \rangle \mathrm{d}u \\
 & - \frac{1}{2} \int_0^T \int_E \int_E f(u, x) f(u, y) \mathcal{Q}(\mu(u); \mathrm{d}x, \mathrm{d}y) \mathrm{d}u \\
 \leq & \int_0^T \|\dot{\mu}(u) - A^*(\mu(u))\|_{\mu(u)} \mathrm{d}u,
 \end{aligned}$$

which, together with Theorem 3.5, implies that

$$I_\nu(\mu(\cdot)) \leq \int_0^T \|\dot{\mu}(u) - A^*(\mu(u))\|_{\mu(u)} \mathrm{d}u. \tag{A.1}$$

We next show that

$$I_\nu(\mu(\cdot)) \geq \int_0^T \|\dot{\mu}(u) - A^*(\mu(u))\|_{\mu(u)} \mathrm{d}u. \tag{A.2}$$

Since v is non-degenerate and $\mu(\cdot)$ is in \mathcal{H}_v , we have that

$$I_v(\mu(\cdot)) = S_v(\mu(\cdot)) < \infty.$$

For any $[s, t] \subset [0, T]$, and any f in $C^{1,0}([s, t] \times E)$, let

$$I_{s,t}(f) = \langle \mu(t), f(t) \rangle - \langle \mu(s), f(s) \rangle - \int_s^t \left\langle \mu(u), \left(\frac{\partial}{\partial u} + A \right) f \right\rangle du.$$

Let

$$L^2([s, t] \times E) = \left\{ f: \int_s^t \int_E f^2(u, x) \mu(u, dx) du < \infty \right\},$$

and L is the linear subspace of $L^2([s, t] \times E)$ of all functions which are constant in space variable x . Let $L^2([s, t] \times E)/L$ be the quotient space of $L^2([s, t] \times E)$ modulo L . We introduce on $L^2([s, t] \times E)/L$ the following norm:

$$\|h\| = \left\{ \int_s^t \int_E \int_E h(u, x) h(u, y) Q(\mu(u); dx, dy) du \right\}^{1/2},$$

and the inner product

$$\langle h_1, h_2 \rangle = \int_s^t \int_E \int_E h_1(u, x) h_2(u, y) Q(\mu(u); dx, dy) du.$$

It is not hard to check that space $(L^2([s, t] \times E)/L, \|\cdot\|)$ is a pre-Hilbert space. Let $\mathcal{L}^2([s, t] \times E)$ be the completion of space $L^2([s, t] \times E)/L$ (cf. for existence refer to p. 56 of Yosida, 1980). Then $(\mathcal{L}^2([s, t] \times E), \|\cdot\|)$ becomes a Hilbert space.

Let $\mathcal{L}_{\text{sub}}^2([s, t] \times E)$ denote the closure in $\mathcal{L}^2([s, t] \times E)$ of the linear span of the set $C^{1,0}([s, t] \times E)$. By an argument similar to Dawson and Gärtner (1987) on pp. 277–280, and the Hahn–Banach Extension Theorem (cf. p. 106 of Yosida, 1980) we have that $I_{s,t}$ is a bounded linear functional on $\mathcal{L}_{\text{sub}}^2([s, t] \times E)$. Thus by the Riesz Representation Theorem, one can find a function h in $\mathcal{L}_{\text{sub}}^2([0, T] \times E)$ such that

$$I_{s,t}(f) = \int_s^t \int_E \int_E f(u, x) h(u, y) Q(\mu(u); dx, dy) du, \quad (\text{A.3})$$

$$\inf_{f \in C^{1,0}([s, t] \times E)} \int_s^t \int_E \int_E (h(u, x) - f(u, x))(h(u, y) - f(u, y)) \times Q(\mu(u); dx, dy) du = 0, \quad (\text{A.4})$$

$$S_v(\mu(\cdot)) = \frac{1}{2} \int_0^T \int_E \int_E h(u, x) h(u, y) Q(\mu(u); dx, dy) du, \quad (\text{A.5})$$

which implies that

$$\begin{aligned} & \int_0^T \|\dot{\mu}(u) - A^*(\mu(u))\|_{\mu(u)} du \\ &= \int_0^T \sup_{g \in C^\infty(E)} \left\{ \langle \dot{\mu}(u) - A^*, g \rangle - \frac{1}{2} \int_E \int_E g(x) g(y) Q(\mu(u); dx, dy) \right\} \end{aligned}$$

$$\begin{aligned}
&= \int_0^T \sup_{f \in C^{1,0}([0,T] \times E)} \left\{ \langle \dot{\mu}(u) - A^*, f(u) \rangle - \langle \mu(u), \dot{f}(u) \rangle \right. \\
&\quad \left. - \left\{ \frac{1}{2} \int_E \int_E f(u, x) f(u, y) Q(\mu(u); dx, dy) \right\} \right\} \\
&= S_v(\mu(\cdot)) - \frac{1}{2} \inf_{f \in C^{1,0}([0,T] \times E)} \int_0^T \int_E \int_E (h(u, x) - f(u, x))(h(u, y) - f(u, y)) \\
&\quad \times Q(\mu(u); dx, dy) du \\
&= S_v(\mu(\cdot)),
\end{aligned}$$

where Eq. (A.3) is used in deriving the third equality, and Eqs. (A.4) and (A.5) are used in reaching the last equality. This together with Theorem 3.5 implies Eq. (3.22). \square

Note added in proof

We have recently improved Theorem 2.5 to $M_1^\tau(E)$ and believe that Theorem 3.4 can be improved in the same way. We also believe that Theorem 3.7 can be improved to a full LDP in the τ -topology.

References

- Bauer, H., 1981. *Probability Theory and Elements of Measure Theory*. Academic Press, London.
- Dawson, D.A., 1978. Geostochastic calculus. *Canadian J. Statist.* 6, 143–168.
- Dawson, D.A., Gärtner, J., 1987. Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. *Stochastics* 20, 247–308.
- Dembo, A., Zeitouni, O., 1993. *Large Deviations and Applications*. Jones and Bartlett Publishers, Boston.
- Donsker, M.D., Varadhan, S.R.S., 1975. Asymptotic evaluation of certain Markov process expectations for large time, I. *Comm. Pure Appl. Math.* 28, 1–47.
- Ethier, S.N., Kurtz, T.G., 1993. Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* 31, (2), 345–386.
- Ethier, S.N., Kurtz, T.G., 1994. Convergence to Fleming–Viot processes in the weak atomic topology. *Stoch. Proc. Appl.* 54, 1–27.
- Freidlin, M.I., Wentzell, A.D., 1984. *Random Perturbations of Dynamical Systems*. Springer, New York.
- Hartl, D.L., Clark, A.G., 1989. *Principles of Population Genetics*, 2nd ed. Sinauer Associates, Inc., Sunderland, MA.
- Pukhalskii, A.A., 1991. On functional principle of large deviations. In: Sazonov, V., Shervashidze, T. (Eds.), *New Trends in Probability and Statistics*. VSP Moks’las, Moskva.
- Shiga, T., 1981. Diffusion processes in population genetics. *J. Math. Kyoto Univ.* 21, 133–151.
- Shiga, T., 1990. A stochastic equation based on a Poisson system for a class of measure-valued diffusion processes. *J. Math. Kyoto Univ.* 30, 245–279.
- Ströock, D.W., Varadhan, S.R.S., 1979. *Multidimensional Diffusion Processes*. Springer, Berlin.
- Wright, S., 1949. Adaptation and selection. In: Jepsen, G.L. (Ed.), *Genetics, Paleontology, and Evolution*. Princeton University Press, Princeton, pp. 365–389.
- Yosida, K., 1980. *Functional Analysis*, 6th ed. Springer, Berlin.